DOCUMENT RESUME

ED 245 791 PS 014 300

AUTHOR Bernal, Ernest M.
TITLE Assessing Language Minority Students in the New Follow Through.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [81]
CONTRACT NIE-P-80-0162
NOTE 42p.
PUB TYPE Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; Bilingual Education; Bilingual Education Programs; Early Childhood Education; *Language Proficiency; Language Tests; Limited English Speaking; Models; Non English Speaking; *Research Needs; State of the Art Reviews; *Student Evaluation; Teacher Evaluation; Test Bias; *Test Interpretation; *Test Validity
IDENTIFIERS New Follow Through Program; Project Follow Through

ABSTRACT

Focusing on the assessment needs of language minority students in the early elementary years and on the evaluation of programs servicing them, this discussion directs specific attention toward accommodating language minority students in the New Follow Through Program. Introductory remarks offer recommendations for developing New Follow Through models for culturally and linguistically integrated settings and for developing tests for English-proficient and limited-English-proficiency children. The first major section describes the state of the art in assessing language minority students. Several ways tests are misused are pointed out and language proficiency assessment, testing school achievement, and teacher assessments are discussed. The second major section examines variables thought to be important in describing programs for language minority students and in studying the relationships of such programs to various student characteristics and local conditions. Model, program, classroom, and student variables are specified and discussed in terms of problems associated with instrumentation and measurement and with respect to measuring variables of interest. The final section identifies problem areas associated with the evaluation of bilingual programs. It is concluded that the inclusion of language minority students in the New Follow Through Program poses challenges and opportunities for curricular, psychometric, and evaluative innovation. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PS

ED245791

ASSESSING LANGUAGE MINORITY STUDENTS IN THE NEW FOLLOW THROUGH

by

Ernest M. Bernal, Ph.D.
Creative Educational Enterprises, Inc.

PS 014300

Prepared under Order No. NIE-P-80-0162
National Institute of Education

Neither NIE nor the Department of Education is responsible for the content or opinions expressed herein; no endorsement of same is implied nor should be inferred.

Introduction and General Comments

This paper focuses on the assessment needs of language minority students in the early elementary years and on the evaluation of programs which service them. If, as proposed, language minority students are to play a significant role in the New Follow Through Program (New FT), then the importance of addressing their measurement looms large indeed, for without methodological refinement in instrument design and selection most of the achievement and affective data collected will probably be worthless, much will undoubtedly be suspect, and only a little will clearly merit the cost of analysis.

Today the importance of properly assessing bilingual students is recognized because of their growing numbers, geographical dispersion, the influence of numerous federal and state programs specifically designed to meet their needs, the impact of federal court cases, and the voice of the bilingual educational constituency. It is becoming profitable to produce tests for this market, and a number of tests have been developed to measure a variety of language and achievement constructs for bilingual or potentially bilingual children. Unfortunately, the quality of most of these instruments, particularly instruments which measure aspects of language proficiency, leaves much to be desired (Bernal, 1977).

Psychometrics as a field is reluctantly becoming aware of the challenges to test validity posed by children for whom tests in English alone have so far inadequately assessed their aptitudes, attitudes, achievement, and development. Obviously, many bilingual children can be tested appropriately with extant English instruments. The problem is that it is difficult to tell who these are ahead of time without conducting other assessments. The point is that language minority populations are culturally-linguistically different,

and this means that they are behaviorally different, often even in the realm of test-taking behaviors.

Similarly, many popular evaluation schemes, such as those which require pre-posttesting with all-English achievement batteries, may be thoroughly confounded by apparent gains made by certain students in social studies or science when what has really happened is that they have learned to read. At the lower ranges, students who remain limited in English proficiency (LEP) experience a cumulative deficit, since norms wait for no one. Should the scores of these two groups be averaged together... presto! No gains. Evaluation designs need to be especially sensitive to crucial intervening variables and, as will be shown later, to special conditions which affect studies in naturalistic settings.

How bilingual children may be accommodated in the New FT is also an important question. In the "old" Follow Through study, bilingual children were so much extraneous "noise" for most of the models and did not figure prominently in the analyses. Under the New FT, they will be included from the outset, but whether language minority children are relegated to certain New FT models or are accommodated by all models has important implications for evaluation.

One way to accommodate language minority students is for all of the New FT models to make provision for some type of parallel bilingual instruction. While this might sound preposterous to some, consider that models must be adaptable to a great number of school sites, that LEP children are found in virtually all major school systems and many small ones, that several states have mandatory "bilingual" education for LEP students, and that desegregation efforts could impact a program by introducing LEP children into a previously English monolingual New FT setting.

Another way to proceed would be to designate two--and preferably more--models as bilingual models under the New FT. This would be a step in keeping with now traditional thought in compensatory education. Yet even here some of the concerns just discussed may have bearing. Would such models be used exclusively with language minority children? If so, how might they accommodate desegregation orders? Bilingual education has not always fared well in desegregated settings (see Zirkel, 1977). If non-LEP, non-language minority children can participate with LEP and even with bilingually proficient children, would the model provide a diluted bilingual treatment, one which would implicitly give preference to English speakers and seek to "transition," or "reclassify," or "exit" the language minority child at the earliest possible moment? This might happen, for instance, if the bilingual model were really nothing more than a set of English-as-a-second language activities appended to an English model.

This writer feels that any exclusively defined bilingual model would be limited in its applicability to ethnically stable, relatively homogenous sites characterized by a large majority of language minority students. Many educators would, of course, find such a model useful, assuming that it would be effective. Furthermore, the measurement and evaluation issues alluded to earlier could be isolated in the New FT impact study. But such a model would not answer the needs of most children or most educators at most sites, would not be in keeping with the original thrust of Follow Through, which Hodges (1978) characterized as a genial innovation "for tying theory and research to actual educational practice" (p. 190). It would not, in short, educe our best efforts to innovate, to anticipate the future, and to make the best use of current knowledge in curriculum development, testing, and evaluation.

Perhaps the most innovative thing which the New FT could do for

compensatory education would be to produce interventions which do not appear to be compensatory, do not egregiously press children to learn and become "just like normal Americans." Perhaps we should start thinking about producing quality educational programs (González, 1979), complex curricula which provide different options for children--strands, if you would. Perhaps we need to devise tests which approach the challenge of measurement not from the perspective of convenience for the test maker or the test user but with a view to being able validly to assess many populations--including the dominant Anglo cultural group--simultaneously or in equivalent ways.

And so I come to the end of these introductory remarks with a recommendation: The National Institute of Education (NIE) should prepare RFPs for two feasibility studies: (1) developing New FT models for culturally and linguistically integrated settings, (2) devising or adapting tests in the cognitive, achievement and affective domains of interest to the New FT for English proficient and LEP children. Alternatively, NIE could prepare non-binding RFPs on either or both of these topics to see if feasible, defensible proposals turn up.

We need to know what our options are, realistically. If we do not consider deliberately the relationship of language minority students to the entire New FT effort, their presence by design or accident may become a nuisance, a "noise" or cacophony which our interventions, instruments, and methodology are ill prepared to orchestrate.

Assessing Language Minority Students: The State of the Art

Briefly...

Hispanics and other language minority groups have become victims of test abuse and test misuse because (1) they have not been adequately represented in the samples of students used for test development (Green, 1972), (2) their language characteristics and lack of test sophistication have not been taken into account in research and evaluation designs or in individual test interpretation and educational decision making, (3) test results have too often been of little practical value, and (4) staff knowledge of test scores has produced a self-fulfilling prophecy effect in school settings (De Avila & Havassy, 1974). For example, whereas IQ and related tests have served to misdiagnose disproportionately large numbers of Hispanic children into mentally retarded or language and learning disability categories (Gerry, 1973), these instruments have not been especially helpful in identifying children at the other end of the ability spectrum, the gifted (Bernal, 1974).

Although a few testing companies have in recent years been making progress in meeting some of these testing problems and developing more valid tests for minority groups, psychologists in the field of measurement and test developers have generally not dealt with these issues and have not sought to impact those aspects of test misuse which are under their control or influence (Bernal, 1975). Instead, those that have articulated on the issues have either shifted the blame to the practitioner (e.g., Cleary, et al, 1975) or, arguing that tests have sufficient validity for some purposes (often predictive validity), have been satisfied to indicate that test scores merely describe the parameters of the problem, but do not create it

(e.g., Jacobson, 1977).

Still, legal and social pressures and a haunting, if vague dissatisfaction with a seemingly endless litany of apologies has caused test developers and psychometrists to take steps to rectify abuses and misuses in the field. Unfortunately, the measures undertaken have frequently been the source of new problems while not really ameliorating the basic condition.

Malpractices, In Passing

The first malpractice, most often found in field settings, consists of "adding points" to obtain scores of language minority students. This procedure is, of course, basically a way of making low test scores more palpable, since it does nothing to increase a test's validity. Sometimes the number of points to be added is subjectively but experientially determined; in other instances the number is based on the average difference between Anglo and minority scores--a very questionable practice indeed, especially when applied to individuals. The method is wrong but the motive for adding points is that educators working with minority children sometimes find that many of them have achieved more than the test scores indicate. Doubtlessly, one of the reasons why various national and state educational organizations have not been friendly to the use of certain types of tests, especially with minority populations, is that too many teachers do not believe their results (e.g., NEA, ca. 1980).

A second malpractice involves simple renorming, i.e., the computation of ethnic norms, often locally. Renorming accomplishes what adding points does, but the numbers are determined empirically. The only real advantage of renorming is that it provides good descriptive statistics for a particular ethnic population and a better distribution of scores. But renorming appears to the uninitiated to do more, to somehow make the test better. It does not.

Test translation without tryout and subsequent modification and validation has also become a popular practice, whether done by a testing company or locally by a practitioner. Sometimes only the directions are translated, but often the entire test is recast into another language, usually Spanish. I have witnessed individually administered tests presented in both languages, a procedure which involves the repetition of each item and which produces an unsystematic practice effect on scores, depending on a child's bilingual skills.

Some testing companies' brochures illustrate English and translated versions of a test in a way which suggests that they are parallel forms, when in fact no empirical verification or equating procedure has been attempted, not even back translation, a technique which has proven so useful in equating the meanings of statements in cross cultural research (Manaster & Havighurst, 1972). In fact, some translated, multiple choice tests are so "parallel" that even the position of the correct answer is unchanged--a measurement travesty when one considers that both versions are sometimes administered to the same students in quick succession--again an untoward practice effect. Furthermore, some translated tests have no norms for the non-English language version; test users are left to assume that the English norms are applicable.

The psychometric and practical problems with test translation are many. Obviously some types of tests, such as simple psychomotor or discrimination tasks or straightforward computation problems, can usually be presented in another language with little adaptation, particularly so when no reading is required of the examinee. Even here, however, cultural content should be checked and test directions back translated, whenever appropriate, and submitted to a trial phase. Vocabulary tests or problem solving tasks involving

cultural content or internal verbal mediation cannot be simply translated without risking the alteration of item characteristics or the factor structure of the tests. In other words, translation usually changes the difficulty range of an item (e.g., if spangle is translated to lentejuela, the item changes in difficulty for Hispanic students). Translation may also change the options a student may otherwise have in answering an item (e.g., stamp may be a verb or a noun in English, but timbre, estampilla, or sellar in Spanish limit the usage of the word). Items recast into another language may be more or less useful in differentiating more accomplished students from their below average peers. Finally, a test which measures one factor for Anglos (e.g., practical intelligence: "What should you do if you cut your finger?") might be measuring another factor for Hispanics (e.g., degree of acculturation to Anglo values and practices), especially if scoring criteria have a limited range of acceptable responses.

Most often translated tests use a relatively formal standard dialect to produce expeditiously a test which will appeal to as wide a group of potential customers as possible. The result, tragically, is that some language minority students who speak a dialect of the language and who have not had sufficient bilingual education, score low on tests in both languages. In still other cases (fortunately few) all language minority children entering school for the first time are tested exclusively in the non-English language, thereby penalizing those who are most proficient in English, a special case of test misuse which once again places language minority students in a disadvantaged situation.

Another malpractice is the administration of selected subscales of larger diagnostic and intelligence tests to language minority students. If this practice were based on empirical findings of greater reliability or

validity for certain subtests, there would be little reason to object; however, this practice usually rests on the belief that LEP students score higher (i.e., look better) on some subscales than on others. Performance subscales, for example, are often preferred by practitioners over verbal scales, in spite of the fact that basing general interpretations on performance tests has usually yielded disappointing results, both for the Anglo population (Nunnally, 1959) and for different cultural and national groups as well (Anastasi, 1976). As a rule, then, the decision to administer only certain subtests to language minority students should be based on empirical studies which incorporate relevant linguistic and ethnographic variables in their designs.

The last malpractice to be discussed is the profligate use of so-called out-of-level testing with LEP children. The argument goes among some psychometrists and evaluators that since LEP children modally score so low on English-based achievement tests, some technique is necessary to generate more variance and normalize the distributions. Out-of-level testing does this--ostensibly--but makes interpretation difficult even with the application of expanded standard scores. Such testing, in my opinion, is rarely used to enhance individual diagnosis. Instead, these data are summarized, and the resultant reports often becloud the problem--however lamely--with passing references to the normative standard or the introduction of grade equivalent explanations. Out-of-level testing, in short, becomes a statistical legerdemain for "adding points."

All of these malpractices have come about because of one simple fact, often intuited but rarely admitted and important to the New FT's planning: there are precious few reliable, valid tests to use with LEP students. The prescription is also simple, or at least straightforward: develop tests--

from scratch, where necessary--which adequately measure constructs of interest in these populations.

Language Proficiency Assessment

The practice of testing non-English native language skills has become so tied to bilingual education, so intertwined with entry-classification and exit-reclassification practices, which involve the additional assessment of English language skills, that it is difficult to speak about the assessment of proficiency in one language alone and without thinking its use in educational decision making. The problems and issues which besiege the testing of non-English language skills--and the opportunities to improve the associated measurement strategies and instruments--are closely paralleled in the English domain, and since we will be mainly discussing the testing of children who are actually or nascently bilingual, it is easy to meld one's concerns and thoughts on these subjects.

It seems also that the testing of non-English native language skills is in most bilingual program settings not done as often or as extensively (in terms of different aspects of language) as the assessment of English language skills. In this author's opinion this occurs primarily because extant federal and state program eligibility requirements emphasize English over native language skills. Then, too, the fact that most bilingual programs are transitional in nature probably augers for greater testing in English, instruction in which is closer to traditional educational ideology (Banks, 1979).

So although research on bilingualism in the schools indicates the importance of measuring both languages on a regular basis, the testing of Spanish language skills continues to languish in both quality and frequency

(Bernal, 1977). The time has come for intelligent, multidisciplinary cooperation among bilingual educators, linguists, and psychometricians to design and produce a variety of valid instruments (Bernal, Note 1) which simultaneously address the pedagogical and classificatory needs of bilingual programs and evaluators' and researchers' needs for versa ile and accurate measures of educationally consequential skills.

Historically the concepts of limited English speaking ability (LESA) and more recently limited English proficiency (LEP) were established more for compliance accountability than for curricular planning. LESA can be measured by tests which measure a student's aural comprehension and speaking proficiency in English; most extant tests of language proficiency were developed at the time when this construct was in vogue. LEP, on the other hand, is more comprenensive, at least beyond the second grade or so, when skills in reading and writing ascend in importance. To this writer's knowledge, no single test or test battery for measuring LEP beyond the second grade exists, and unless new Title VII regulations or Lau guidelines specify or operationalize this construct, individual programs must determine what this means for themselves (Bernal, Note 2).

Many bilingual programs, unfortunately, administer only the English parts of these language assessment tests, reasoning--in this case correctly--that a LESA child is also LEP. In the later elementary grades, however, a non-LESA child may be LEP, as discussed above. Several good reviews of Spanish proficiency tests exist (e.g., Silverman, Noa, & Russell, 1976; Dieterich, Freeman, & Crandall, Note 3), but they seem to converge weakly on the conclusion that few good ones are to be had, and are more useful for indicating what to avoid than for what to do.

Still, knowledge of LEP status alone, without data of the child's

ability in Spanish, has limited usefulness for designing appropriate interventions. Given a LEP child entering school for the first time, information about her/his Spanish competencies might lead us to suspect the validity of the testing administration, help us decide to refer the child for further assessment, provide important placement information, or screen children whose native or English language skills might be very "mature," i.e., much more developed than those of their typical agemates.

Cummins (1979) suggests that the continuous assessment of bilingual students' progress in the development of native language skills is important, particularly if one wants to predict their success in an all-English educational environment. Native language achievement is an indicator of students' general academic potential in English as well. His very recent theoretical work (Cummins, 1980) distinguishes between basic interpersonal communication skills (BICS) and cognitive/academic language proficiency (CALP), the metalinguistic skills which provide the kinds of learning advantages that some bilinguals seem to enjoy. It is the CALP in one's first language (L1) which predicts success in the second language (L2) environment. It seems that extant language proficiency tests mostly measure BICS, and thus have limited utility for diagnosis or educational placement and classification.

Language proficiency utilizes criterion-referenced, norm-referenced, or a combination of both techniques to establish the level of an examinee's language mastery, and it can be measured through interview techniques or paper-and-pencil tests, depending on the aspects of language (productive or receptive skills) one wishes to define as appropriate to a particular age/grade level or to a specified role/situation (such as the proficiency required for a bilingual teacher). Tests of language proficiency, unlike popular measures of vocabulary and reading, emphasize aspects of linguistic

competence.

Language dominance is a construct properly reserved for the nascent or functioning bilingual. It may be defined operationally as the higher of two language proficiency levels. There is a great demand for measures of language dominance, particularly for Hispanics, from early childhood through the early elementary years. Bilingual education and ESL programs variously use language dominance appraisals to accept children, place them in instructional groupings, assess their language progress, evaluate certain aspects of curricula, and in the case of transitional bilingual education, to determine the appropriate point at which a student is ready to exit the bilingual program and enter the English monolingual course of instruction ordinarily offered in the schools.

Language dominance assessments made without an examination of language proficiency have, in my opinion, fostered two related and tacitly held beliefs which desensitize educators to individual differences. One is that children cannot be proficient in the language in which they are not dominant; the other is that children must be competent in their dominant language. Some bilingual children--like some monolinguals--do have a language dysfunction, and this affects their language competence even in their dominant language. Normal and, certainly, gifted children acquire two language systems readily, although they may still be more proficient in one of them.

The quality of tests for measuring language proficiency varies considerably, but not one to date is truly outstanding, or even satisfactory. This author has served on an advisory committee on proficiency assessment to the Texas Education Agency.* The committee has reviewed dozens of

*Committee for the Evaluation of Language Assessment Instruments (CELAI).

tests and has found them all wanting to a greater or lesser degree in terms of traditional psychometric criteria or linguistic content and organization (CELAI, Note 4).

In their worst forms language proficiency tests pander to schools' tight bilingual budgets and some districts' desires not to identify LEP children, while others rest on highly questionable assumptions, purport to measure the impossible (e.g., purport to measure language dominance without measuring language proficiency), or do not guide the users whatsoever in the interpretation of the results. The rest too often present linguistically unrealistic demands (e.g., "Now we're going to talk in English," or "Please use complete sentences") or arbitrary scoring or weighting procedures, and generally suffer from a lack of sensible items, sufficient language sampling, and reliability. Then, too, not even a handful have been validated against groups of proficient monolinguals, and none have been examined in the light of the demand characteristics of bilingual or English monolingual classrooms. Tests that use scales or ordered categories of proficiency are not sensitive to the student who is marginally proficient in either language alone but nevertheless communicatively competent in informal bilingual settings where codeswitching is the rule. The fact that an increasing number of bilingual programs are including non-LEP (i.e., English proficient, or EP) students in the program, for many of whom Spanish is in effect a second language, also necessitates the testing of Spanish language skills.

Testing School Achievement

The testing of school achievement areas in a non-English language poses other problems, not the least of which is the lack of tests well suited for many language minority groups. A few tests in Spanish are on

the market, but these are essentially translated--and in some cases renormed --versions of the English achievement series published by the same company. These translations have item-by-item similarity with the original tests in English, which effectively precludes the use of both English and Spanish versions of the same level test on the same bilingual children, since exposure to one would produce a practice effect for the other. A notable exception is the CIRCO test battery (Bernal, Note 5); Hardy, Note 6), which was adapted (not merely translated) from the CIRCUS (in English) and which includes new tests developed for Hispanic students. This test, however, can be used only with four to six year olds. Most non-Hispanic, language minority groups simply have not had standardized achievement tests developed for them to date.

This and the fact that locally developed instruments do not have the credibility of commercial instruments are the principal reasons why achievement testing is conducted by and large in English. Such testing of LEP children, however, produces considerable personal and statistical fallout.

In this writer's evaluation consultations with bilingual programs in public school settings, he has seen instances where grade level averages, involving dozens of bilingual classes in several schools, have just reached chance level performance on nationally standardized achievement tests. I cannot say, of course, just how widespread a phenomenon this is, but if OBEMLA's* plans to implement (on a voluntary basis) a standard Title VII data reporting form (Baca, Bernal, DeGeorge, & Mangino, Note 7) are carried out, then such data can be calculated.

These tests can be frustrating to LEP children, and their results, as discussed earlier, often lack credibility with teachers. For evaluators,

*Office of Bilingual Education and Minority Languages Affairs.

too, the results can be frustrating, since a large percentage of scores at or below chance makes for fairly arbitrary interpretations of the results, interpretations which historically have placed the burden on minority children, the inadequacy of their educational programs, and their economic circumstances, instead of calling for a reexamination of the tests themselves (Bernal, 1975).

Recent test reviews of English-based tests used in bilingual programs (e.g., Silverman, Noa, & Russell, 1976) have started to emphasize test appropriateness as an important aspect of validity. The valid application of a test assumes that the examinees are not unlike the group(s) upon which the test was developed and standardized. To the extent that important psychological differences exist (such as in cultural background and language proficiency), test results must be interpreted with caution and supplemental measures of the trait or construct in question should be utilized to cross-check the results.

In achievement testing, too, a psychometric lag occurs, since we know how to obtain "better" performance from language minority students on standardized tests (Bernal, 1977), performance which increases their scores and enhances test reliability and so, hopefully, measurement validity. Specific recommendations will be included in a later section of this paper.

Teacher Assessments

In addition to the assessment of students there is a growing trend to assess teachers in the non-English (and sometimes in English) language skills (Carlisle-Zepeda, & Saldate, 1978). This author endorses this trend (although he is aware of the political agendas which sometimes motivate it), because one reason why so many bilingually certified teachers do not teach

in Spanish is that their basic Spanish skills are inadequate or that their content-related vocabulary is lacking. Effective teachers in bilingual classrooms are both professionally and interpersonally articulate (Rodríguez, 1980).

One of the techniques being used to assess teachers' non-English language competency is the Language Proficiency Interview (LPI), a taped and blindly scored version of the technique used by the Foreign Service Institute and the Peace Corps to test the conversational abilities of their trainees. The LPI is used by New Jersey, and Texas uses it in addition to a standardized, multiple-choice test of Spanish language competencies for previously certified teachers seeking to obtain a bilingual endorsement through additional course work. Obviously cloze tests and other procedures could be used. What is still needed is a better measure of teachers' non-English writing skills. Techniques for estimating the adequacy of writing samples in English have been developed, and these could be adopted for the assessment of non-English writing skills as well.

Summary

Bilingual assessment is now recognized as crucially important for the selection of bilingual teachers, for the screening, placement, and reclassification of LEP students, and for designing appropriate bilingual educational programs. Extant instruments, unfortunately, are not entirely capable of satisfying these needs.

What Should Be Measured to Study Program Effects on Language Minority Students in the New Follow Through: An R&D Agenda

In this section we will examine variables which are particularly important for describing programs for language minority students and studying the interactions of these programs with various student characteristics and local conditions. It is assumed that many other variables will be considered in the ordinary progression of events in evaluation design and implementation, so emphasis will be placed here on bilingual models, students, and sites.

Model/Program/Classroom Variables

The following table (Table 1) presents in summary form the model/program/classroom variables of high potential interest to the New FT.

Table 1

Model/Program/Classroom Variables for Studying the Impact of the New Follow Through on Language Minority Students

Teacher language proficiency in English:

- Speaking proficiency: conversational, general educational, and curricular areas.
- Writing skills, general.

Teacher proficiency in the non-English language:

- Speaking proficiency: conversational, general educational, and curricular areas.
- Writing skills, general (when applicable).*

*Some non-English languages found in bilingual programs have no standard written form.

Table 1 - continued

Instructional aide's language proficiency in English:

- Speaking proficiency: conversational and curricular areas.
- Writing skills, general (when applicable).*

Instructional aide's proficiency in the non-English language:

- Speaking proficiency: conversational and curricular areas.
- Writing skills, general (when applicable).*

Division of instructional duties in L1 and L2 by teacher and aide.

Instruction given in L1 and L2:

- Total time in each.
- Percent of instruction in each.
- Content areas affected by each: reading, math, social studies, etc.

Formal and informal language interactions between aide and teacher, teacher or aide and students:

- Use of L1 and L2 by function (e.g., instruction, encouragement, direction, discipline).

Parental participation:

- In parent-teacher conferences.
- In bilingual classroom activities.
- In parent training (if applicable).
- Use of L1 and L2 in these activities.

Instructional Management:

- "Pull-out" vs. integral.
- Timing of L2 introduction: delayed, simultaneous, or immersion.

Degree of ethnic/linguistic integration:

- Ethnic/language minority groups represented.
- Language proficiency categories represented.

While the instructor's proficiency in English and the non-English language has always been regarded by experts as crucial for the success of bilingual education programs (for example, Center for Applied Linguistics, 1974; California State Department of Education, Note 8), recent empirical

evidence (Rodríguez, 1980) tends to show that higher levels of fluency and linguistic flexibility differentiate the better bilingual educators from their average and below average colleagues. Numerous approaches to the measurement of instructor proficiency have been tried (Bernal, Note 2), ranging from evidence of college credit in the non-English language to structured interviews and standardized paper and pencil tests.

Most of these techniques, however, have been "make-do," and none singly satisfies the need to measure all receptive and productive language skills. This writer's hunch is that the implementation of the non-English facets of a bilingual program depend in no small way not only upon teachers' and aides' conventional skills, but also upon their ability to converse professionally in the language, on their knowledge of content-related phrases, and on their general writing ability. The same may be said for their skills in English. Bilingual educators, in other words, probably need to be competent to understand, speak, read, write, and teach the languages involved in a particular program model before they will actually use them to any great extent in the classroom or for communicating with language minority parents.

In order to measure these skills adequately, some techniques need to be applied to the task. The New FT should call for the adaptation of the Language Proficiency Interview (LPI) (ETS, Note 9) to include discussions about professional topics and instructional areas. The LPI technique records these interviews and has them scored blindly by trained raters who use a criterion referenced scale with five or six major and four or five minor ranks. If correctly adapted, such a scheme could yield separate scores for general and pedagogical areas.

Although writing skills in the non-English language could be measured by multiple-choice tests, this writer urges that more demanding tests be

developed on already existing techniques. Dictation and cloze test techniques (Oller, 1975; Oller & Streiff, 1975) are particularly good, inexpensive techniques to measure the grammatical components of writing skills. Cloze tests, furthermore, can be cast into multiple-choice formats. Scaling of these techniques probably needs additional attention, however.

Writing samples scored holistically can also be employed, although this procedure requires somewhat more expensive scoring by teams of readers. Identifying and assembling groups of highly competent readers for certain languages could be difficult, but major languages such as Spanish or French could be handled readily in this manner. The advantages of writing samples centers around their face validity, is that the examinee must produce an essay on one or more assigned topics designed to permit scope of expression. Fairly reliable scoring techniques have been developed for English writing samples, and the same could be done for other written languages.

The division of instructional duties in L1 and L2 by teacher and aide has been of concern to bilingual educators for some time. Many of them are basically concerned that the more prestigious person--the teacher--will conduct instruction in English and relegate the use of the non-English language to the aide, thereby influencing language attitudes in an undesirable manner (Bernal, Note 10). But there are additional issues which have to do with a program model's implementation in a classroom and, very importantly, with time-on-task, which seems to be particularly important for compensatory educational programs (Davidson & Holley, 1979).

Thus instruction given in L1 and L2 involves total time, percent of instruction in each language, and the content areas affected by these instructional modes. Understanding who teaches what in which language and for how long is potentially important for teasing out differences in program impacts (Saville-Troike, 1978).

Instructional logs or questionnaires could be designed by the model developers to keep track of these figures. But occasional spot checks would be necessary to validate these procedures, and it is hoped that the evaluation of the New FT incorporates plans to conduct extensive on-site observations.

Which brings us to the next topic, formal and informal language interactions between aide and teacher, teacher or aide and students. Classroom observation procedures such as those inspired by Flanders (1961) could be designed for use by bilingual observers. A short term developmental effort is foreseen in this area to test the observational system's usability and reliability, keeping in mind that similar procedures have required careful observer training, spot checking (Reid & DeMaster, 1972), and recalibration (Reid, 1970). Adding a bilingual dimension to such systems may require compromising the scope of the interactions to be observed or the use of an additional observation schedule which focuses on other interactive processes. It remains to be seen whether an effective bilingual interaction form may be used to supplement an observation technique already well developed.

Parental participation data can probably be supplied accurately by program personnel by keeping good records of meetings and other types of contacts. During the recent conference on the Longitudinal Evaluation of Bilingual Programs (see Bernal, 1980b; Contreras, 1980), the positive and negative effects of different kinds and levels of parental participation become evident. The New FT program should monitor these effects carefully.

Many aspects of instructional management could be emphasized in the New FT. Two general concerns arising from practices in the field (Bernal, Note 10) have been selected for inclusion, since data on these elements should tell us much about New FT programs' approaches to teaching language minority children, especially LEP children.

The first concern has to do with classifying a program as "pull-out" or integral. The chief characteristic of the pull-out approach is that non-English instruction is provided only to LEP students and conducted by a resource teacher who works with them for a limited period during the day. It is not unusual in such a program for the "bilingual" teacher to service several organized classes by working with small groups in a learning center in the "home room," or to conduct a number of special classes made up of students "pulled out" of their regular classrooms for bilingual instruction or tutoring in English.

An integral program provides bilingual instruction to children in a regular classroom setting by the regular classroom teaching staff. Academic content is to some extent taught in both languages, and often non-LEP children participate in these activities as well.

Obviously, some classrooms may be "mixed," as in the case where sufficient instructional resources exist for one language minority group but not for another.

The second concern has to do with timing the introduction of English to the LEP child. The delayed introduction of English for instruction in the content areas requires that such instruction be essentially monolingual in the non-English language and that the study of English assume the status of a subject in a broader curriculum. As competencies in English are acquired by the children, the academic areas assume a more bilingual orientation. Children in such programs are often able to read and write in their first language before they are introduced to these skills in English.

The simultaneous introductions of English begins content instruction in both languages from the first day. Severely LEP students may be provided some additional help in the native language, but by and large the atmosphere in decidedly bilingual and one language helps to support the other. English

reading and writing instruction are introduced as quickly as possible.

An immersion approach is difficult to characterize accurately. It is not an English-as-a-second-language (ESL) technique, and it is not the traditional "sink-or-swim" system which LEP students have so long endured. It is, instead, a carefully sequenced system of content instruction which helps children intuit the language as classes progress (Cohen, 1975; Lambert & Tucker, 1972; Bernal, Note 10). Learning aids, highly animated teaching, and sensitive adjustments of the English demand characteristics of the classes educe the desired English skills, ostensibly without jeopordizing the normal development of the mother tongue. Non-English instruction is later introduced as a subject.

The degree of ethnic/linguistic integration will require the use of detailed demographic instruments down to the classroom level. Accurate language proficiency categories will have to be devised, since current instruments do not deal effectively with all bilinguals and since other categorical systems, e.g., the Lau categories (Hai, 1978), would seem to be better adapted to legal and administrative classification than for scientific inquiry.

It is difficult to say which categories linguists and educators will agree are of interest to a New FT evaluation study. This writer believes that two types of categorical systems should be used. The first is based on the LEP-EP distinction. EP students can be divided into dominant English-speaking minority and language minority populations, and the latter can be subdivided into once-LEP (reclassified) and never-LEP students. This scheme would account for the presence of Anglo, other English-speaking minority populations (principally Black students), and English monolingual language minority background students. It would also identify those language minority

children who are currently LEP, those who were once LEP and are now reclassified as EP (and hence important to follow up), and those who were functionally bilingual when they first entered the program. The presence or absence of EP language minority students will indicate whether the program is being implemented in a transitional mode, i.e., whether it exits students who become EP and does not directly service those language minority students who come to school already competent in the target language.

The second type of language-based categorical system advocated here is based upon a three-dimensional matrix of functional English and non-English language categories and communication competence. A child's placement in this system would depend not only upon her/his relative performance in each of two languages, but also upon their ability to cope with a variety of language tasks. Two-dimensional categorizations (based on English and non-English languages), it is recognized, already exist in the literature on bilingual proficiency assessment and in certain program regulations. What is envisioned here, however, is a system which is capable of better diagnostic-prescriptive applications (particularly for students who score at the lower ranges of both English and non-English scales), is not misled by the spontaneous (and, one might add, highly adaptive) codeswitching behavior exhibited by some bilingual children, incorporates current language analysis theory, and measures that aspect of language development (CALP) which predicts readiness to engage in second language instruction. This will be discussed further in the next section.

Both of these systems should be seen as dynamic, rather than static. Categorical membership and changes in categorical membership can be seen, respectively, as important covariables for the study of program-by-student interactions, or as criteria for program effectiveness.

Student Variables

Given the state of the art in language proficiency assessment, good tests using a meaningful metric must be developed through a coordinated, multidisciplinary R&D effort. The experience gained during the development of CIRCO (Bernal, 1977; Bernal, Note 5; Hardy, Note 6) suggests to this writer that an English proficiency test can be constructed to accurately measure the status of several language minority populations, so long as great care is taken to reduce disabling test anxiety (Sarason, 1961) and to prepare them for the testing experience. The content of such an instrument, furthermore, should be established on native English speakers of the same age, so that no items be included for language minority students that English speaking members of the dominant ethnic group cannot themselves pass. Basing performance on native English speakers, in fact, could be one way of establishing a meaningful metric for English proficiency.

Lest we start thinking of this development effort only in traditional terms, let me hasten to say that linguists have some innovative ideas for judging the level of language development, including some incisive techniques to analyze mistakes and the child's differential use of both languages. The principal shortcoming of these procedures, by psychometric standards, is their inefficiency. This is why a multidisciplinary effort seems particularly appropriate at this juncture (Bernal, Notes 1 & 11).

The contemplated test should measure LEP in the more comprehensive sense previously explicated. This means essentially that students in the second or third grades must be tested in English reading and writing in addition to oral language proficiency. The determination of content should pose little problem since the more popular commercial achievement tests would seem to have sampled these curricular domains quite well. Indeed the SWRL* Student

*Southwest Regional Laboratory for Educational Research and Development.

Placement System (SWRL, 1980), intended to assist in the assessment and placement of language minority students, appears to have largely duplicated the efforts of commercial testing companies (Potter, Note 12) at the early elementary grade levels.

The innovations required for a test of LEP, in this writer's opinion, pivot around techniques for (1) screening children for eligibility for the English test, and (2) accommodating their diverse expectancies and test-taking behaviors. The screening procedure envisioned would be a brief, painless, and valid way of categorizing LEP children at the lower ranges, children who should not be exposed to a longer, frustrating examination in a language they barely understand. CIRCO has shown that a brief test in Spanish can be used to select and operationally a group of students for whom its Spanish-based subtests are appropriate (Bernal, 1977). There is no reason to believe that a similar process could not be used in English assessment of LEP--or in the administration of English-based general achievement test batteries, for that matter.

Other writers (see Bernal, 1977) have used techniques for reducing untoward test anxiety, enhancing motivation, and familiarizing students with those demand characteristics of the test which are not central to the measurement objective but which if misunderstood could cause students to receive lower marks than they would otherwise achieve, i.e., would introduce systematic error into their measurement. Such techniques, as argued elsewhere in this paper, have not received sufficient attention from psychometricians, yet are pivotal to testing language minority students and minority populations in general (Bernal, 1975).

The other half of the language assessment picture is the measurement of non-English proficiency, and particularly of cognitive/academic language proficiency (CALP). Now CALP as a construct is at the cutting edge of

theories of bilingualism, so it may be difficult to operationalize. We do know some things about it, however, including that it seems to be measured best by discrete-point (i.e., decontextualized) items of higher-order cognitive processes mediated by the native language. Verbal learning psychologists need to examine CALP along with psycholinguists to see how similar it seems to be to such cognitive mechanisms as verbal mediation. If CALP turns out to be closely related to factors which are psychometrically more familiar, then instrument design can move ahead relatively quickly, although, of course, it may have to be cast in several languages.

This requirement for producing diverse tests of non-English language proficiency poses a potential financial issue for the New FT. Designing and developing different tests in a systematic way for Spanish, French, Navajo, Chinese, Vietnamese, and other language minority groups would be an expensive proposition. Consequently a recommendation is in order. The New FT should commission the development of (1) a comprehensive and broadly comprehensible test of LEP, (2) one or two tests of non-English language proficiency and achievement according to anticipated need, and (3) a compatible general technique for testing the non-English proficiency and achievement of other participating language minority groups. Under the second part of this recommendation, proficiency tests in Spanish and perhaps one other language would be developed on a priority basis. The general technique espoused in the third part of this recommendation might be developed around guidelines for criterion-referenced measurement of the relevant language domains.

Were the New FT to decide to measure only LEP status to the exclusion of non-English language proficiency, an important diagnostic and classificatory base would be lost. CALP is too exciting, too potentially useful a construct to overlook. Were the New FT to restrict the number of different

language minority groups participating in the New FT program, it would have to either restrict the types of program sites to those which could introduce no "surprises" in the evaluation design or find suitable techniques for converting the highly probable statistical "noise" into orchestratable patterns. The only other alternative is to be willing to sacrifice important empirical data to the gods of finance.

Achievement testing in the non-English language is another matter. There are many reasons for promoting subject matter achievement testing in L1 for LEP students, but none, I believe, should put the burden of supporting their development on the New FT. In this area of measurement extant English-based achievement tests can be made to suffice so long as adequate safeguards can be developed to protect LEP children from test misuse.

These safeguards could include the use of the comprehensive test of English proficiency for screening. Assuming that such a test would provide valid assessments, there would be little point in subjecting profoundly LEP children to a four hour battery in English. But there may be some need to investigate several related issues further: (1) what should be the cutoff point on the comprehensive English proficiency test for excusing students from the achievement test; (2) are there any parts of standard achievement batteries which can be administered validly to LEP children with or without minor adaptation (i.e., adaptations which do not jeopardize the comparability of scores)? Similarly, children in the New FT--whatever their ethnicity--should receive practice in test-taking skills as part of any model's curriculum, thereby enhancing the children's competence to cope with such instruments (Saville-Troike, 1978).

The "old" FT used other cognitive measures in addition to achievement. If similar plans are being made as of this writing, this author would like

to suggest the use of Piagetian tests as measures of cognitive maturity which many language minority populations can take without in their own languages significant bias (De Avila & Havassy, 1974). These tests also have diagnostic possibilities for special programs (De Avila & Havassy, 1975), including the identification of gifted children (Bernal, 1974).

Similarly, a judicious sample of classrooms might be administered a test of cognitive style. The literature on the relevance of cognitive style to instructional effectiveness and teacher-student relations establishes the importance of this variable for education (Witkin, Note 13) and cross-cultural research (Witkin, 1967). Then, too, instructional techniques may have differential effects on language mincrity children with different cognitive styles (Holtzman, Goldsmith, & Barrera, 1979).

In the affective domain, attitude measurements should include the esteem which language minority and majority populations have for each other and the attitudes of the language minority group to the use of their language and toward their own ethnic group. Monitoring these attitudes, particularly as language minority children grow in their English proficiency, should be one way of estimating some of the programmatic effects of concern to language minority populations and rounding out the evaluation of the New FT models.

In this section we have discussed the need for a major R&D effort to develop an adequate, multiculturally appropriate test of English language proficiency, at least one test of proficiency in a non-English language, and a complementary general technique for testing the native language proficiency of numerically small ethnic groups for which quality, standardized assessments are not likely to become available. So far as subject matter achievement testing of language minority populations is concerned, the

better extant standardized instruments (both norm-referenced and criterion-referenced) can be made to suffice so long as LEP children are not placed at risk. Piagetian measures of intellectual development and tests of cognitive style round out the cognitive domain. In the affective domain interethnic and language attitudes should be included in the New FT's plans to evaluate programmatic effects.

Cautions in Evaluation

The discussions about instrumentation, measurement and bilingual education in this paper have implications for evaluation in the New FT. This paper is not devoted to evaluation, but since the evaluation of bilingual programs is fraught with difficulties, as evidenced by the AIR study (Danoff, 1978) and its aftermath (see, for example, O'Malley, Note 14), a few problem areas will be identified.

One has already been mentioned, the preponderance of chance scores. O'Malley (Note 14) noted in his review of the AIR data that even when averages seemed to favor the children in the bilingual programs, they were rarely higher than the 20th percentile on national norms. It is clear, then, that data collected in compensatory programs are often highly skewed positively, and that significant proportions of language minority students score at or below chance on multiple-choice tests. If some of the suggestions for protecting LEP students and finding alternative achievement measures made in this paper are followed, more useable data should result.

Another caution which needs to be observed is the imposition of unreasonable standards of performance on LEP children. A second language is not acquired like skills in an academic subject. In the past language minority students are seen as making considerable improvements in English and in tested academic achievement only in the later elementary grades (USGAO, 1976; O'Malley, Note 14). Cummins' (1980) work suggests that CALP takes time to develop, and that if it doesn't develop in the native language it may never develop at all. Since FT has limited itself in the past to the early elementary years, it may not be possible to show massive growth in English language proficiency and academic achievement without a followup study.

Placing LEP children in non-bilingual comparison classes is ethnically questionable and often unfeasible (Bissell, 1980). In some states, furthermore, it is illegal. This writer's experience indicates that in public school settings the exigencies of teaching the children most in need make a shambles out of randomization efforts. Designs which take advantage of intersite, interclassroom variability in student characteristics and instructional approaches should be used, since these may prove more useful than using models as independent variables (House, et al., 1978; Rodríguez-Brown, 1978).

Lack of process and contextual data restricts the interpretation of efforts. Obtaining data on program characteristics is crucial (Rodríguez-Brown, 1978), and such variables have been recommended herein. Ethnographic monitoring (Hymes, 1979) should also be considered in a sample of sites, since this may gain data from another perspective not only on processes and contexts but also on effects, especially on unanticipated outcomes.

High student attrition can be expected to occur in the New FT among language minority students generally, if the Title VII experience is any indicator (see Ligon, 1980).

> ...for many schools large attrition rates indicate... unsystematic "exits"... due to the exigency of serving the students most in need with limited resources or the recalcitrance of some local school administrators who would subotage the program by convincing the parents of moderately well achieving students to sign waivers because their children presumably "don't... need the program anymore." The cumulative effect of these practices is probably to depress the average scores of the remaining project students.... (Bernal, 1980a).

Special cautions and agreements between the New FT and participating schools are in order, else student cohorts may be capriciously dismembered. Large numbers of students and classrooms should be obtained whenever possible.

Conclusions

The inclusion of language minority students in the New FT poses great challenges and opportunities for curricular, psychometric, and evaluative innovation. This paper, in delineating variables of interest to the New FT and the means of measuring them, has hopefully disabused us of any facile notion that merely including these students and setting aside an instructional model or two for them will suffice. The New FT will doubtlessly have to accommodate language minority students in ways never envisioned in the 1960s. These challenges should be met creatively, not just expeditiously, in the tradition of Follow Through, which is to bring the best of educational theory into the realm of educational practice.

Reference Notes

1. Bernal, E. M. What makes language test an expletive: Tests, test selection, and test use. In D. G. Larson (Chair), What makes test a four letter word? Viewpoints on bias in standardized testing. Symposium at the annual meeting of the Teachers of English to Speakers of Other Languages, Boston, February-March 1979.

2. Bernal, E. M. Problematic applications of languag proficiency testing in bilingual education. In D. A. Turner and V. Flores (Co-Chairs), ESL in bilingual education: Implementation and other practical considerations. Colloquium presented at the annual meeting of the Teachers of English to Speakers of Other Languages, Boston, February-March 1979.

3. Dieterich, T. G., Freeman, C., & Crandall, J. A. A linguistic analysis of some English proficiency tests. Paper presented at the meeting of the National Association for Bilingual Education, Seattle, May 1979.

4. Committee for the Evaluation of Language Assessment Instruments, Texas Education Agency. Report of the Committee, winter and spring 1979. Unpublished report, Division of Bilingual Education, TEA, 1979.

5. Bernal, E. M. CIRCO: An innovative response to assessment needs of young Spanish-speaking children. Paper presented at the Third National Conference on Multicultural Curriculum and Materials, San Francisco, February 1976.

6. Hardy, R. Should José be tested in Spanish? A validation study for the El Circo test series. Paper presented at the annual meeting of the National Council on Measurement in Education, April 1977.

7. Baca, R., Bernal, E. M., DeGeorge, G. P., & Mangino, E. Title VII basic grant and demonstration program evaluation reporting form and report outline. Form submitted for publication, 1981.

8. California State Department of Education. Criteria for bilingual teacher competencies. Report available from California SDE, Sacramento CA. Dated July 1977.

9. Educational Testing Service. Language proficiency interview. Unpublished manuscript, ETS, Princeton NJ, undated.

10. Bernal, E. M. Dimensions of bilingual education: A look at the real world. Paper presented at the National Institute of Education Conference on Dimensions of Bilingual Programs, Washington DC, February 1977.

11. Bernal, E. M. Testing Spanish native language skills in bilingual education. Paper presented at the Summer Symposium on Bilingual Multicultural Education Research, Buffalo NY, May-June 1980.

12. Potter, L. Report of the CAL/NCBE review panel of the SWRL student placement system. Unpublished report, Center for Applied Linguistics, Washington DC, September 1980.

13. Witkin, H. A. The role of cognitive style in academic performance and in teacher-student relations. Paper presented at the Symposium on Cognitive Styles, Creativity, and Higher Education, Graduate Record Examination Board, Montreal, November 1972.

14. O'Malley, J. M. A re-examination of the evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Unpublished manuscript, 1979. (Available from the National Institute of Education, Washington DC)

References

Banks, J. A. Multiethnic/multicultural teacher education: Conceptual, historical, and ideological issues. Rosslyn VA: National Clearinghouse for Bilingual Education, 1979.

Bissell, J. S. Program impact evaluations: An introduction for managers of Title VII projects. Austin TX: Evaluation, Dissemination and Assessment Center for Bilingual Education, 1980.

Bernal, E. M. Gifted Mexican American children: An ethnoscientific perspective. California Journal of Educational Research, 1974, 261-273.

Bernal, E. M. A response to "educational uses of tests with disadvantaged subjects." American Psychologist, 1975, 30, 93-95.

Bernal, E. M. Assessment procedures for Chicano children: The sad state of the art. Aztlán, 1977, 8, 69-81.

Bernal, E. M. Evaluation of curriculum development: Discussion. In The executive summary of the Conference on Longitudinal Evaluation of Bilingual Programs. Austin TX: Austin I.S.D., Office of Research & Evaluation, 1980(a).

Bernal, E. M. Longitudinal evaluation of bilingual programs. National Clearinghouse for Bilingual Education Forum, November 1980(b), pp. 1,4.

Carlisle-Zepeda, V., & Saldate, M., IV. And who assesses the bilingual teacher's language proficiency? Educational Leadership, 1978, 35 318-321.

Center for Applied Linguistics. Guidelines for the preparation and certification of teachers of bilingual-bicultural education. Arlington VA: CAL, 1974.

Cleary, T. A., Humphreys, L., Kendrick, S., & Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.

Cohen, A. D. A sociolinguistic approach to bilingual education: Experiments in the American southwest. Rowley MA: Newbury House, 1975.

Contreras, R. Parental involvement: State of the art in Austin ISD Title VII bilingual education project, 1974-80. In The executive summary of the Conference on Longitudinal Evaluation of Bilingual Programs. Austin TX: Austin ISD, Office of Research & Evaluation, 1980.

Cummins, J. Linguistic interdependence and the educational development of bilingual children. Review of Educational Research, 1979, 49, 222-251.

Cummins, J. The entry and exit fallacy in bilingual education. National Association for Bilingual Education Journal, 1980, 4.

Danoff, M. N. Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Overview of study and findings. Palo Alto CA: American Institutes for Research, 1978. (ERIC Document Reproduction Service No. ED 154 634)

Davidson, J. L., & Holley, F. M. Your students might be spending only half of the school day receiving instruction. American School Board Journal, March 1979, pp. 40-41.

De Avila, E. A., & Havassy, B. IQ tests and minority children. Austin TX: Dissemination Center for Bilingual Bicultural Education, 1974.

De Avila, E. A., & Havassy, B. Piagetian alternatives to IQ: Mexican American study. In N. Hobbs (Ed.), Issues in the classification of exceptional children. San Francisco: Jossey-Bass, 1975.

Flanders, N. A. Analyzing teacher behavior as part of the teaching-learning process. Educational Leadership, December 1961, 19, pp. 173-180.

Gerry, M. H. Cultural myopia: The need for a corrective lens. Journal of School Psychology, 1973, 11, 307-315.

González, J. M. Towards quality in bilingual education; bilingual education in the integrated school. Rosslyn VA: National Clearinghouse for Bilingual Education, 1979.

Green, D. R. Racial and ethnic bias in test construction. Monterey CA: CTB/McGraw-Hill, 1972.

Hodges, W. L. The worth of the Follow Through experience. Harvard Educational Review, 1978, 48, 186-192.

Hai, T. T. The Lau categories revisited. IDRA Newsletter, December 1978, pp. 1-2, 4-7.

Holtzman, E. H., Goldsmith, R. P., & Barrera, C. Field-dependence field-independence: Educational implications for bilingual education. Austin TX: Dissemination and Assessment Center for Bilingual Education, 1979.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No simple answers: Critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48, 128-160.

Hymes, D. Ethnographic monitoring. In E. J. Briere (Ed.), Language development in a bilingual setting. Los Angeles: National Dissemination and Assessment Center, 1979.

Jacobson, R. L. Standardized testing and cultural bias. Chronicle of Higher Education, July 25, 1977, pp. 3-4.

Lambert, W. E., & Tucker, G. R. Bilingual education of children: The St. experiment. Rowley MA: Newbury House, 1972.

Ligon, G. D. Technical issues encountered in the five-year Title VII evaluation. In The executive summary of the Conference on Longitudinal Evaluation of Bilingual Programs. Austin TX: Austin ISD, Office of Research & Evaluation, 1980.

Manaster, G. J., & Havighurst, R. J. Cross-national research: Social-psychological methods and problems. New York: Houghton Mifflin Co., 1972.

National Education Association. NEA focus: Standardized testing, the frustrating system that traps our students. Washington DC: NEA, ca. 1980.

Oller, J. W. Cloze, discourse, and approximations to English. In M. K. Burt and H. C. Dulay (Eds.), New directions in second language learning, teaching, and bilingual education. Washington DC: Teachers of English to Speakers of Other Languages, 1975.

Oller, J. W., & Streiff, V. Dictation: A test of grammar based expectancies. In R. L. Jones and B. Spolsky (Eds.), Testing language proficiency. Arlington VA: Center for Applied Linguistics, 1975.

Reid, J. B. Reliability assessment of observation data: A possible methodological problem. Child Development, 1970, 41, 1143-1150.

Reid, J., & DeMaster, B. The efficacy of the spot-check procedure in maintaining the reliability of data collected by observers in quasi-national settings: Two pilot studies. Oregon Research Institute Research Bulletin, 1972, 12, No. 8.

Rodríguez, A. M. Empirically defining competencies for effective bilingual teachers: A preliminary study. Bilingual Education Paper Series (Vol. 3, No. 12). Los Angeles: National Dissemination and Assessment Center, 1980.

Rodríguez-Brown, F. V. The do's and don'ts in regard to the evaluation of bilingual programs. Paper presented at the annual meeting of the National Association for Bilingual Education, San Juan PR, April 1978. (ERIC Document Reproduction Service No. ED 156 717)

Sarason, S. B. Anxiety and learning. In W. B. Waetjen (Ed.), Human variability and learning: Papers and reports from the Fifth Curriculum Research Institute. Washington DC: Association for Supervision and Curriculum Development, National Education Association, 1961.

Saville-Troike, M. A guide to culture in the classroom. Rosslyn VA: National Clearinghouse for Bilingual Education, 1978.

Silverman, R. J., Noa, J. K., & Russell, R. H. Oral language tests for bilingual students: An evaluation of language dominance and proficiency instruments. Portland OR: Northwest Regional Educational Laboratory, 1976.

Southwest Regional Laboratory for Educational Research and Development. Resources for developing a student placement system for bilingual programs. Los Alamitos CA: SWRL, 1980.

United States General Accounting Office. Bilingual education: An unmet need (USGAO Publication No. MWD-76-25). Washington DC: USGAO, 1976.

Zirkel, P. A. Bilingual education and school desegregation: A case of uncoordinated remedies. Bilingual Review, 1977, 4, 180-188.